Signal Processing Challenges in Climate Data; Global Temperature and CO_2

David J. Thomson

AT&T Bell Laboratories, Murray Hill, N.J. 07974

Abstract

Possible changes in the world's climate resulting from human use of fossil fuel is perhaps the most serious problem facing the human race. The statistical problems in this area are also challenging; are we changing the climate in a measurable way, or are the currently perceived changes simply a result of natural variability? This paper outlines some of the *statistical* problems and a few examples that can be found in climate statistics.

Introduction

The major problem in the statistical analysis of climate data is to determine whether currently perceived changes in the Earth's climate are being caused by anthropogenic changes in the atmosphere, the "greenhouse effect," from burning fossil fuels or, alternatively, if these changes could plausibly be the result of naturally occurring changes[1, 2] such as variations in the Earth's orbit, internal oscillations[3], or of the Sun's output. There are (at least) two major statistical problems: *First,* the time-scale of *known* natural variations in climate is immense, ranging from the 5-minute solar *p*-modes, through the 208 and 2400 year changes in solar activity measured in ^{14}C variations and tree-ring thicknesses, to the 2.9 million year periodic terms in precession of the Earth's axis. *Second,* the structure of many of the time-series describing these processes appears to be more complicated than those we now understand statistically. In the following sections I attempt to summarize a few of these problems with emphasis on parts where I have personal knowledge. The basic methods I use[5 - 9] were summarized at the Adelaide ICASSP[10].

Available Data

The first problem that must be faced in the analysis of climate data is the lack of data. *Instrumental* data is limited to the last few centuries; the earliest such temperature record, that of Central England[11, 12] begins in 1659. Next, much of what data is available consists only of monthly averages. Apart from such averages not coming from a proper low-pass filtering and resampling operation, one must contend with the unequal lengths of the months and possible aliases of the 27 day rotation of the sun, the 29 day synodic lunar month, and probably others. Almost none of this data was collected for scientific purposes, but primarily for meteorology. As such, little heed was given to discontinuities caused by moving stations (for example from city centers to airports), changing instrumentation, recording times during the day and the like. While new instruments are usually better (some early records in the north have winter "average" temperatures of $-40\,^{\circ}C$ because mercury thermometers froze), even today, when the need for quality control of such data is obvious, new instruments are introduced without allowing a proper (several year) overlap to ensure continuity of effective calibration. (As a specific example of these problems, many of the supposed temperature records broken in the US during the past few years reflect nothing more than that new instruments with faster response times were introduced without proper cross-calibrations.) In addition, even when stations have not been moved, their environment has frequently been changed by urbanization. Many records have gaps from fires, wars, logs lost at sea, etc. Finally, in addition to missing data there are numerous outliers, both from natural causes such as volcanos, and from the usual transcription errors and the like. In both cases, robust methods[13] must be used, but the conclusions should differ.

Because the instrumental data records are so short, much effort has been spent developing "proxy" records. These are natural records that preserve some aspect of the climate or related variables. As a specific example, the departures from a simple exponential decay in ^{14}C concentration in wood as a substitute for direct solar cosmic-ray data are better-than-average because nearly absolute dating is available by counting tree-rings. However, even the dating process, known as "dendrochronology," needs careful signal processing because long series depend on matching overlapping records taken from different trees. Typically, annual variations are small, ring thicknesses in many well-preserved samples are only a fraction of a millimeter, errors in adjacent rings negatively correlated, and, judging from personal experience[6], the underlying process is both non-stationary, and probably has a mixed spectrum. Unfortunately, in many samples, annual ring thicknesses

0-7803-2431-5/95 $4.00 © 1995 IEEE

are too thin to allow accurate ^{14}C determinations, so the records are gappy with some measurements being three or five year "averages." This is an area where some signal processing education would help; a current tendency is to do "precision" ^{14}C measurements, typically with coarse time steps, as opposed to the preferable procedure, when one is unable to prefilter data, of making more measurements on a finer time-step. Thus one must contend with the interpolation problem; the best work seems to be in the astrophysical literature where such problems are the rule; the spectral technique[14] was developed in an attempt to resolve a controversy between the Haskell, see *e.g.* Press[15], and Edelson-Krolic[16] estimators.

The main scientific use of ^{14}C data is as a solar proxy and, like most solar data, has an "unusual" spectrum. (A simple count of papers on the sunspot spectrum is, by itself, reasonable evidence that the process is unusually complicated.) First, the spectrum of ^{14}C variations has a discrete spectrum component[6] at several frequencies but, particularly, the "Suess wiggles" at periods of 104 years, plus multiples, and sub-multiples. In addition, the remainder of the process is non-stationary and taking the singular value decomposition of the dynamic spectrum (as sketched in[6]) shows that the continuous part of the spectrum varies systematically almost certainly with related periodic components. Given the current emphasis in the signal-processing community on higher-order spectra, the irony in the observation that this "logarithmic-order" decomposition conveys much useful information may suggest a new area to study.

A more complex problem arises from assigning reliable dates to longer-term proxy series obtained from geological coring. A typical example in this category are the ^{18}O and ^{13}C records typically obtained by isotopic analysis of fossil foraminifera from ocean-bottom sediment cores. Here the data may extend back several hundred-thousand to a few million years, and the main problem is that samples are obtained as a function of *depth,* not time. Typically, the sedimentation rate is variable, plausibly correlated with climate, and, in addition, disturbance by ocean currents or bioturbation may cause gaps or disordering in parts of cores.

The problem of dating sediment cores has been empirically solved by an iterative process known as "tuning." This process depends on the presence of strong, nearly-periodic components in the data from changes in the obliquity of the Earth's axis on the ecliptic, eccentricity of the orbit, and precession. Tuning begins with an approximate time-scale derived from radiometric dates, geological markers such as volcanic ash-layers, or magnetostratigraphy. Departures in the phase of the filtered data compared to *e.g.* precession are then used to refine the time-scale. The problem with tuning[17] is that, if done on both precession and obliquity the results disagree. The root of the problem appears to be that both precession and obliquity depend on the differences of the moments-of-inertia of the Earth computed about the polar and equatorial axes and these moments-of-inertia depend, in turn, on surface loading. Because full glaciation involves transferring about 10^{19} *kg* of water from the oceans to the polar regions (reducing sea level by about 120 *meters* in the process) the moments-of-inertia change. Thus, although the precession constant is usually given to six decimals, the astronomically determined value differs from that expected from seismic evidence in the third, and it appears that development of an accurate time scale requires understanding of the entire process. The celestial mechanics involved are complicated and Laskar[18] notes the complicating factor of a near resonance in precession. Reconciliation of the astronomically and seismic values of the precession constant, however, may depend on our ability to estimate the transfer functions between glacial loading and the moments-of-inertia. The transfer functions may not be linear and estimates given in[7] differ significantly from those implied by models of ice loading and glacial rebound[19]. Thus a signal-processing challenge: how does one estimate a possibly non-linear[20] transfer function when the time-scale of the observations depends on the transfer function being estimated? Once this is solved there is a "simpler" problem, namely how to tell if the 100,000 year glacial cycle is an approximately linear response to the multiple lines in current eccentricity theories[21] or if a nonlinear alternative theorie, such as[22], is more compatible with the observations. In addition to providing a reasonable baseline for the historical variability of the climate system, one must understand the theory of orbital change to investigate even the instrumental climate record.

Length of the Year

To show the effects of precession in modern temperature, run a monthly temperature record through a narrow-band filter centered at 1 *c/y*. (*c/y* is "cycle per year.") At many locations the phase of the filter output will not be constant but will have a linear drift of about -50 arc-seconds per year. As a specific example, Figure 1 shows the phase of the Central England series[12], obtained from a narrow-band filter centered at 1 *c/y*. The slope is close to the general precession constant, $50''.2/y$ and implies that the frequency of the temperature cycle is approximately one cycle per *anomalistic* year (the time from perihelion-to-perihelion) rather than one cycle per *tropical* year (the time from equinox-to-equinox). Note that even though the tropical and anomalistic years differ by about 1 in 26,000, the frequency offset is crucial.

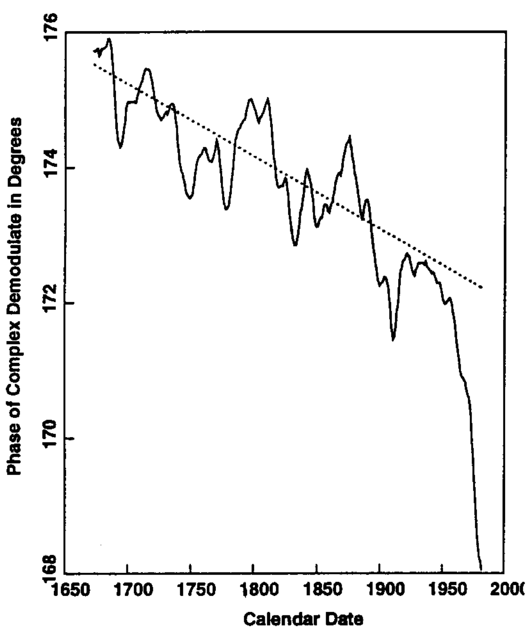

Figure 1: Phase of the Central England Temperature Series at 1 *c/y. Only the initial phase of the dotted line is adjusted, the slope is determined by precession.*

About the middle of this century the trend that had held from 1659 changed, and the phase began to change rapidly. From Figure 1 I infers that, on average, English winters occur about four days later now than they were in 1945. However, if one does a similar computation for a Northern Hemisphere average series[23, 24] one finds that, while the early phase behavior has the slope of $\approx -50''/y$ (as above), after mid-century the phase slope *increases* rapidly, rather than decreasing. Moreover, using long records from different locations, the phase change has been much more variable in the last fifty years than it was in the preceding fifty.

To see why such small frequency offsets are important, consider the process of "deseasonalizing" routinely used to produce temperature anomaly series[24]. (The annual cycle often accounts for 90 percent of the variance in records from temperate latitudes and deseasonalizing is a procedure designed to remove what was thought to be a large, stable, systematic, and generally uninteresting component from the raw data.) Consider a true temperature variation $A\cos(2\pi t - \theta - \phi(t))$ where $\phi(t)$ is a small phase departure from nominal, and subtract a seasonal cycle assuming the frequency to be $1\ c/y$, $A\cos(2\pi t - \theta)$. For small $\phi(t)$, simple identities give a residual $A\ \phi(t)\sin(2\pi t - \theta)$ so when $\phi(t)$ has a drift, as shown above, the anomaly series will have a seasonally dependent trend. The magnitude of this effect can be surprisingly large; in polar regions the amplitude of the annual cycle can be as much as 32 °*C*, so a 50 arc-second per year offset gives a seasonal change in slope of ± 0.78 °*C* per century. By comparison, the average trend in temperature is only about 0.65 °*C* per century, so the error is immense. A glance at Figure 1, moreover, shows that the rate of phase change has altered dramatically since mid-century, so current seasonal trend estimates may have errors five or more times that given above.

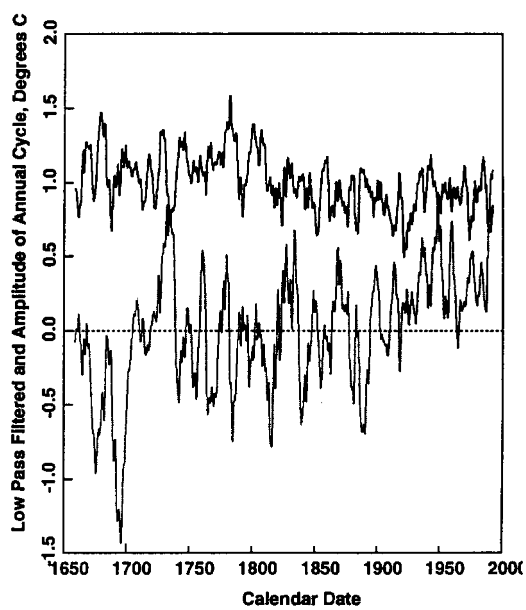

Figure 2: Low-Pass (lower) and Annual Cycle Amplitude for the Central England Series. The amplitude has been offset.

Solar Variability

Of the various alternatives to greenhouse gas forcing for the temperature changes observed during the last century, the most plausible is that advanced by Friis-Christensen and Lassen[4] who showed that, if the *period* of the sunspot numbers were used as a proxy for solar output, the correlation with average temperature was excellent. This, or any other, hypothesis of solar variability can be tested by the simple observation that changes in the solar "constant" directly modulate both the average temperature and the amplitude of the annual cycle so there should be coherent variation between them. Figure 2 shows both the low-frequency component, $[0, .5\ c/y)$ and a projection filter(25) demodulation of the annual cycle, $(0.5, 1.5\ c/y)$ and, although the trend in the low-frequencies is positive, the amplitude of the annual cycle has been constant to slightly decreasing since about 1850. At several frequencies generally associated with solar activity the two signals are reasonably coherent, and using a multiple-window coherency calculation[26] gives the results shown in Figure 3.

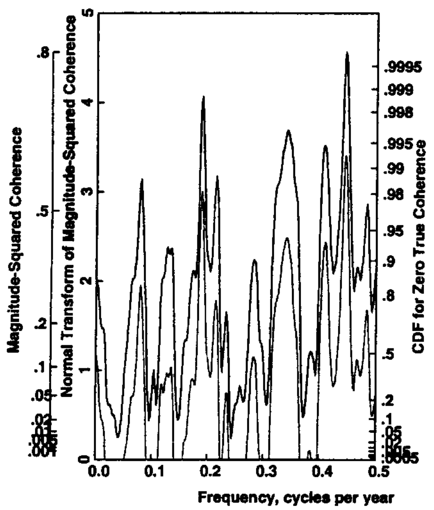

Figure 3: Coherence between Low-Pass and Amplitude. The outer left scale gives MSC, inner left a normal transform, and the right, the probability of observing the given level with independent series. The lower trace is one jackknife standard deviation below the estimate.

The phase, not shown, is 180 ° at low frequencies and otherwise shows about a two-year delay, possibly the transport time of the Gulf Stream. While there are feedbacks from average temperature to the amplitude of the seasonal cycle[28], they should not be large enough to cancel solar effects completely, so the small negative coherence observed at low frequencies effectively eliminates solar variability as a cause of the observed warming.

References

(1) Lindzen, R. S., *Bull. Amer. Meteor. Soc.* **71,** 288-99 (1990).

(2) Schneider, S. H., *Science* **263,** 341-7 (1994).

(3) Mann, M. E. & J. Park, "Global-scale modes of surface temperature variability on interannual to century timescales," *J. Geophys. Res.* in press, (1994).

(4) Friis-Christensen, E. & K. Lassen, *Science* **254,** 698-700 (1991).

(5) Thomson, D. J., *Proc. IEEE* **70,** 1055-96 (1982).

(6) Thomson, D. J., *Phil. Trans. R. Soc. Lond. A.* **330,** 601-16 (1990).

(7) Thomson, D. J., *Phil. Trans. R. Soc. Lond. A.* **332,** 539-597 (1990).

(8) Thomson, D. J., *Proc. SPIE* **2027,** 236-244, (1993).

(9) Thomson, D. J. & A. D. Chave, Ch. 2 of **Advances in Spectrum Analysis,** S. Haykin, Ed., Prentice-Hall (1991).

(10) Thomson, D. J., *Proc. ICASSP* **VI,** 185-94 (1994).

(11) Manley, G., *Q. J. R. Meteorol. Soc.* **100,** 389-405 (1974).

(12) Parker, D. E., T. P. Legg, & C. K. Folland, *Int. J. of Climatology* **12,** 317-42 (1992).

(13) Kleiner, B., R. D. Martin, & D. J. Thomson, *J. Royal Statist. Soc.* **B-41,** 313-51 (1979).

(14) Thomson, D.J., and R. Schild, *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics, Lake Tahoe, CA, 374-8 (1993).*

(15) *Press, W., Rybicki, G., & Hewitt, J., Ap. J.* **385,** *404-415 (1992).*

(16) *Edelson, R. A., and J. H. Krolik, Ap. J.* **333,** *646-59 (1988).*

(17) *Shackleton, N. J., A. Berger, & W. R. Peltier, Trans. R. Soc. Edin.: Earth Sciences* **81,** 251-261 (1990).

(18) Laskar, J., F. Joutel, & F. Boudin, *Astron. and Astrophysics* **270,** *522-533 (1993).*

(19) *Peltier, W. R., Science* **265,** *195-210 (1994).*

(20) *Imbrie, J. I. et al. Paleoceanography* **8,** *699-735 (1993).*

(21) *Berger, A., & M. F. Loutre, Quaternary Sci. Rev.* **10,** 297-317 (1991).

(22) Ghil, M., *Physica* **D 77,** 130-159 (1994).

(23) Jones, P. D., "Hemispheric Surface Air Temperature Variations: A Reanalysis and an Update to 1993," *J. Climate,* in press (1994)

(24) Hansen, J. & S. Lebedeff, *J. Geophys. Res.* **92,** 13,345-13,372, (1987).

(25) Thomson, D. J., *Proc. IEEE Seventh SP Workshop on Statistical Signal and Array Processing,* 39-42 (1994).

(26) Kuo, C., C. Lindberg, & D. J. Thomson, *Nature* **343,** 709-14 (1990). (Reprinted in pp 395-400 of **Coherence and Time Delay Estimation,** G. C. Carter, Ed., IEEE Press, 1993.)

(28) Manabe, S., & R. J. Stouffer, *J. Geophys. Res.* **85,** 5529-5554 (1980).